(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) **EP 3 734 472 A1**

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
04.11.2020 Bulletin 2020/45

(51) Int Cl.:
G06F 16/34 *(2019.01)* ***G06F 40/30*** *(2020.01)*

(21) Application number: **19211879.2**

(22) Date of filing: **27.11.2019**

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR
Designated Extension States:
BA ME
Designated Validation States:
KH MA MD TN

(30) Priority: **28.04.2019 CN 201910349382**

(71) Applicant: **Beijing Xiaomi Intelligent Technology Co., Ltd.**
Beijing 100085 (CN)

(72) Inventors:
- **PAN, Xiaotong**
  Beijing, Beijing 100085 (CN)
- **LIU, Zuopeng**
  Beijing, Beijing 100085 (CN)

(74) Representative: **dompatent von Kreisler Selting Werner -**
Partnerschaft von Patent- und Rechtsanwälten mbB
Deichmannhaus am Dom
Bahnhofsvorplatz 1
50667 Köln (DE)

(54) **METHOD AND DEVICE FOR TEXT PROCESSING**

(57) Provided are a method and device for text processing. The method includes that: word segmentation is performed on a first text to be processed to determine multiple words of the first text; one word is removed from the first text to obtain one second text; each word is removed from the first text one by one to obtain a second text set; the first text and each second text in the second text set are analyzed respectively to obtain a first tree structure of the first text and a second tree structure of each second text; and a summary text of the first text is determined according to the first tree structure and each second tree structure.

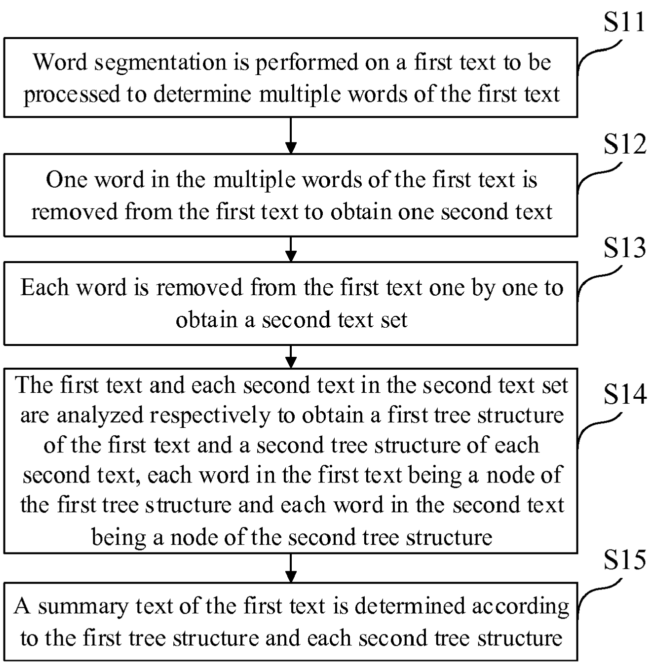

FIG. 1

EP 3 734 472 A1

Description

TECHNICAL FIELD

[0001] The present disclosure generally relates to the technical field of computers, and more particularly, to a method and device for text processing.

BACKGROUND

[0002] In related art, graph ranking, deep learning and the like are usually adopted to extract, from a text, a summary text which is capable of expressing the theme of the text and has a length shorter than a length of the text. According to the graph ranking method, commonly used stop words, prefixes and suffixes or part-of-speech information, etc. may be adopted to filter a redundant text out. However, the graph ranking method have some limitations, for example, key information is easily ignored, information such as a word order and a syntactic structure is not considered, the obtained summary text may be not fluent. For the deep learning method, a deep learning model is needed to be trained based on a large amount of annotation data, and statistical information of words is needed to be determined based on annotations in context to finally determine the summary text, and moreover, in this method, the word order, syntactic structure and the like information are not considered either.

SUMMARY

[0003] For solving the problem in the related art, the present disclosure provides a method and device for text processing.

[0004] According to a first aspect of embodiments of the present disclosure, a method for text processing is provided, which may include that:

word segmentation is performed on a first text to be processed to determine multiple words of the first text;
one word in the multiple words of the first text is removed from the first text to obtain one second text;
each word is removed from the first text one by one to obtain a second text set;
the first text and each second text in the second text set are analyzed respectively to obtain a first tree structure of the first text and a second tree structure of the each second text, each of the multiple words in the first text being a node of the first tree structure and each of words in the second text being a node of the second tree structure; and
a summary text of the first text is determined according to the first tree structure and the each second tree structure.

[0005] In a possible implementation mode, the operation that the first text and the each second text in the second text set are analyzed respectively to obtain the first tree structure of the first text and the second tree structure of each second text may include that:

dependency grammar analysis is made over the first text and the each second text respectively to obtain dependence relationships between the multiple words of the first text and dependence relationships between words of the each second text respectively; and
the first tree structure of the first text and the second tree structure of the each second text are obtained respectively according to the dependence relationships between the multiple words of the first text and the dependence relationships between the words of the each second text.

[0006] In a possible implementation mode, the operation that a summary of the text to be processed is determined according to the first tree structure and the each second tree structure may include that:

scores of structure match between multiple second tree structures and the first tree structure are determined respectively according to the first tree structure and the each second tree structure;
a target tree structure is determined from the each second tree structure according to the scores of structure match of the multiple second tree structures; and
a target word is deleted from the first text to obtain the summary text of the first text, the target word being a word missing from a second text with the target tree structure relative to the first text.

[0007] In a possible implementation mode, the operation that the scores of structure match between the multiple second tree structures and the first tree structure are determined respectively according to the first tree structure and the each second tree structure may include that:

a score of node match between a first target node and a second target node is determined according to the first target node of the first tree structure, a child node of the first target node, the second target node of the second tree structure and a child node of the second target node, the first target node being any word in the first tree structure and the second target node being a word the same as the first target node; and
the scores of structure match between the multiple second tree structures and the first tree structure are determined according to the scores of node match of multiple nodes.

[0008] In a possible implementation mode, the scores of structure match may include scores of structure full match and scores of structure part match.
[0009] According to a second aspect of the embodiments of the present disclosure, a device for text processing is provided, which may include:

a first module, configured to perform word segmentation on a first text to be processed to determine multiple words of the first text;
a second module, configured to remove, from the first text, one word in the multiple words of the first text to obtain one second text;
a third module, configured to remove each word from the first text one by one to obtain a second text set;
a fourth module, configured to analyze the first text and each second text in the second text set respectively to obtain a first tree structure of the first text and a second tree structure of the each second text, each of the multiple words in the first text being a node of the first tree structure and each of words in the second text being a node of the second tree structure; and
a fifth module, configured to determine a summary text of the first text according to the first tree structure and the each second tree structure.

[0010] In a possible implementation mode, the fourth module may include:

a first submodule, configured to make dependency grammar analysis over the first text and the each second text respectively to obtain dependence relationships between the multiple words of the first text and dependence relationships between the words of the each second text respectively; and
a second submodule, configured to obtain the first tree structure of the first text and the second tree structure of the each second text respectively according to the dependence relationships between the multiple words of the first text and the dependence relationships between the words of the each second text.

[0011] In a possible implementation mode, the fifth module may include:

a third submodule, configured to determine scores of structure match between multiple second tree structures and the first tree structure respectively according to the first tree structure and the each second tree structure;
a fourth submodule, configured to determine a target tree structure from the each second tree structure according to the scores of structure match of the multiple second tree structures; and
a fifth submodule, configured to delete a target word from the first text to obtain the summary text of the first text, the target word being a word missing from the second text with the target tree structure relative to the first text.

[0012] In a possible implementation mode, the third submodule may further be configured to:

determine a score of node match between a first target node and a second target node according to the first target node of the first tree structure, a child node of the first target node, the second target node of the second tree structure and a child node of the second target node, the first target node being any word in the first tree structure and the second target node being a word the same as the first target node; and
determine the scores of structure match between the multiple second tree structures and the first tree structure according to the scores of node match of multiple nodes.

[0013] In a possible implementation mode, the score of structure match may include scores of structure full match and scores of structure part match.
[0014] According to a third aspect of the embodiments of the present disclosure, a device for text processing is provided, which is characterized by including:

a processor; and
a memory configured to store an instruction executable for the processor,

wherein the processor may be configured to execute the method of any one of claims 1-5.

[0015] According to a fourth aspect of the embodiments of the present disclosure, a non-transitory computer-readable storage medium is provided, having an instruction therein, when executed by a processor, to enable the processor to execute the method for text processing.

[0016] The technical solutions provided in the embodiments of the present disclosure may have the following beneficial effects: each word in the first text is deleted to obtain the multiple second texts, so that a probability of ignoring key information is reduced; and the first text and each second text are analyzed to obtain the tree structures formed by the words, including word orders and syntactic structures of the first text and the second texts, so that the obtained summary text is semantically fluent.

[0017] It is to be understood that the above general descriptions and detailed descriptions below are only exemplary and explanatory and not intended to limit the present disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

[0018] The accompanying drawings, which are incorporated in and constitute a part of this specification, illustrate embodiments consistent with the present disclosure and, together with the description, serve to explain the principles of the present disclosure.

Fig. 1 is a flow chart illustrating a method for text processing, according to an exemplary embodiment of the present disclosure.

Fig. 2 is a flow chart illustrating the step S13 of a method for text processing, according to an exemplary embodiment of the present disclosure.

Fig. 3 is a schematic diagram illustrating dependence relationships between words of a first text, according to an exemplary embodiment of the present disclosure.

Fig. 4 is a schematic diagram illustrating dependence relationships between words of a second text, according to an exemplary embodiment of the present disclosure.

Fig. 5 is a schematic diagram illustrating a first tree structure, according to an exemplary embodiment of the present disclosure.

Fig. 6 is a schematic diagram illustrating a second tree structure, according to an exemplary embodiment of the present disclosure.

Fig. 7 is a flow chart illustrating the step S14 of a method for text processing, according to an exemplary embodiment of the present disclosure.

Fig. 8 is a schematic diagram illustrating scores of structure match, according to an exemplary embodiment of the present disclosure.

Fig. 9 is a schematic diagram illustrating application of a method for text processing, according to an exemplary embodiment of the present disclosure.

Fig. 10 is a block diagram of a device for text processing, according to an exemplary embodiment of the present disclosure.

Fig. 11 is a block diagram of a device for text processing, according to an exemplary embodiment of the present disclosure.

Fig. 12 is a block diagram of a device for text processing, according to an exemplary embodiment of the present disclosure.

Fig. 13 is a block diagram of a device for text processing, according to an exemplary embodiment of the present disclosure.

DETAILED DESCRIPTION

[0019] Reference will now be made in detail to exemplary embodiments, examples of which are illustrated in the accompanying drawings. The following description refers to the accompanying drawings in which the same numbers in different drawings represent the same or similar elements unless otherwise represented. The implementations set forth in the following description of exemplary embodiments do not represent all implementations consistent with the present disclosure. Instead, they are merely examples of apparatuses and methods consistent with aspects related to the present disclosure as recited in the appended claims.

[0020] Terms used in the present disclosure are only adopted for the purpose of describing specific embodiments but not intended to limit the present disclosure. "A/an", "said" and "the" in a singular form in the present disclosure and the appended claims are also intended to include a plural form, unless other meanings are clearly denoted throughout the present disclosure. It is also to be understood that term "and/or" used in the present disclosure refers to and includes one or any or all possible combinations of multiple associated items that are listed.

[0021] It is to be understood that "first", "second" and similar terms used in the specification and claims of the present application are not to represent any sequence, number or importance but only to distinguish different parts. Likewise, similar terms such as "a/an" also do not represent a number limit but represent "at least one" . Terms like "include" or "contain" refer to that an element or object appearing before "include" or "contain" covers an element or object and equivalent thereof listed after "include" or "contain" and does not exclude another element or object. Similar terms such as "connect" or "interconnect" are not limited to physical or mechanical connection, and may include electrical connection, either direct or indirect.

[0022] The embodiments of the present disclosure will be described below in combination with the accompanying drawings in detail. Characteristics in the following embodiments and implementation modes may be combined without conflicts.

[0023] Fig. 1 is a flow chart illustrating a method for text processing, according to an exemplary embodiment of the present disclosure. As shown in Fig. 1, the method includes the following steps.

[0024] In step S11, word segmentation is performed on a first text to be processed to determine multiple words of the first text.

[0025] In step S12, one word in the multiple words of the first text is removed from the first text to obtain a second text.

[0026] In step S13, each word is removed from the first text one by one to obtain a second text set.

[0027] In step S14, the first text and each second text in the second text set are analyzed respectively to obtain a first tree structure of the first text and a second tree structure of each second text. Each word in the first text is a node of the first tree structure and each word in the second text is a node of the second tree structure.

[0028] In step S15, a summary text of the first text is determined according to the first tree structure and each second tree structure.

[0029] According to the method for text processing of the embodiments of the present disclosure, each word in the first text may be deleted to obtain each second text corresponding to each word of the first text, so that the probability of ignoring key information is reduced. The first text and each second text are analyzed to obtain the tree structures formed by the words, including word orders and syntactic structures of the first text and the second texts, so that the obtained summary text is semantically fluent.

[0030] In a possible implementation mode, the method for text processing may be executed by a terminal device, the terminal device may be User Equipment (UE), a mobile device, a user terminal, a terminal, a cell phone, a cordless phone, a Personal Digital Assistant (PDA), a handheld device, a computing device, a vehicle-mounted device, a wearable device and the like. The method may be implemented by a processor by calling a computer-executable instruction stored in a memory. Or, the terminal device may acquire the first text and send the first text to a server, so that the method is executed by the server.

[0031] In a possible implementation mode, in the step S11, the first text to be processed may be a structure formed by the multiple words such as a paragraph, a sentence, a word group and a phrase. Word segmentation may be performed on the first text according to semanteme of the first text and semanteme of each word of the first text to determine the multiple words of the first text. For example, word segmentation may be performed on the first text through a word segmentation method, such as a maximum matching algorithm, a minimum matching algorithm, a backtracking method and a deep-learning-based word segmentation algorithm, to determine the multiple words of the first text.

[0032] In an example, the first text may be a sentence, for example, "今天北京故宫的天气很好，人很多, (Good weather of the Forbidden City Beijing today, many people)". Any one of the aforesaid word segmentation methods may be adopted for the sentence to determine multiple words of the sentence. In the example, the multiple words of the sentence may be segmented by use of any word segmentation method based on semanteme of each word of the sentence to obtain "今天/北京/故宫/的/天气/很好/，/人/很多(today/beijing/the Forbidden City/of/weather/good/, /people/many)", namely the multiple words of the sentence are "今天 (today)", "北京 (beijing)", "故宫 (the Forbidden City)", "的 (of)", "天气 (weather)",很好 (good)", ",", "人 (people)" and "很多 (many)" respectively.

[0033] In a possible implementation mode, in the step S12, one word may be removed from the first text to obtain a second text, and the obtained second text may correspond to the removed word. For example, the first text is formed by n (n is a positive integer) words, the kth (1≤k≤n) word in the first text may be removed, and the remaining n-1 words in the first text may form a second text corresponding to the kth word, an order of the remaining n-1 words in the first text is kept unchanged. For example, word segmentation is performed on the first text to obtain "今天/北京/故宫/的/天气/很好/，/人/很多 (today/beijing/the Forbidden City/of/weather/good/, /people/many)", and "故宫 (today)" may be removed to obtain a second text

北京/故宫/的/天气/很好/, /人/很多 (beijing/the Forbidden City/of/weather/good/, /people/many)".

[0034] In a possible implementation mode, in the step S13, each word in the first text may be removed one by one to obtain the second text set. For example, the first word in the first text may be removed to obtain the second text corresponding to the first word, the second word in the first text may be removed to obtain the second text corresponding to the second word,..., the nth word in the first text may be removed to obtain the second text corresponding to the nth word. The first text includes n words, and thus n second texts may be obtained, and the n second texts form the second text set.

[0035] In the example, the first text is "今天/北京/故宫/的/天气/很好/, /人/很多 (Good weather of the Forbidden City Beijing today, many people)". After word segmentation is carried out, each word may be removed to obtain multiple second texts. For example, "故宫 (today)" is removed to obtain the second text "北京/故宫/的/天气/很好/, /人/很多 (beijing/the Forbidden City/of/weather/good/, /people/many)"; "的 (beijing)" is removed to obtain a second text "今天/北京/故宫/的/天气/很好/, /人/很多 (today/the Forbidden City/of/weather/good/, /people/many)"; "天气 (the Forbidden City)" is removed to obtain a second text "今天/北京/故宫/的/天气/很好/, /人/很多 (today/beijing/of/weather/good/, /people/many)"; "很好 (of)" is removed to obtain a second text "今天/北京/故宫/的/天气/很好/, /人/很多 (today/beijing/the Forbidden City/weather/good/, /people/many)"; "人 (weather)" is removed to obtain a second text "今天/北京/故宫/的/天气/很好/, /人/很多 (today/beijing/the Forbidden City/good/, /people/many)"; "很多 (good)" is removed to obtain a second text "今天/北京/故宫/的天气/, /人/很多 (today/beijing/the Forbidden City/of/weather/,/people/many)"; "," is removed to obtain a second text "今天/北京/故宫/的/天气/很好/, /人/很多 (today/beijing/the Forbidden City/of/weather/good/people/many)"; "人 (people)" is removed to obtain a second text "今天/北京/故宫/的天气/很好/,/很多 (today/beijing/the Forbidden City/of/weather/good/, many)"; and "很多 (many)" is removed to obtain a second text "今天/北京/故宫/的天气/很好/,/人 (today/beijing/the Forbidden City/of/weather/good/, /people)". The multiple second texts form the second text set.

[0036] In a possible implementation mode, in the step S14, the first text and the multiple second texts may be analyzed respectively, for example, analysis is performed according to semanteme of the first text and the second texts and semanteme of the words therein, and the first tree structure of the first text and the second tree structure of each second text may be obtained; each word in the first text is a node of the first tree structure and each word in the second text is a node of the second tree structure.

[0037] Fig. 2 is a flow chart illustrating the step S14 of a method for text processing, according to an exemplary embodiment of the present disclosure. As shown in Fig. 2, the step S14 includes the following steps.

[0038] In the step S141, dependency grammar analysis is made over the first text and each second text respectively to obtain dependence relationships between the multiple words of the first text and dependence relationships between words of each second text respectively.

[0039] In the step S142, a first tree structure of the first text and a second tree structure of each second text are obtained respectively according to the dependence relationships between the multiple words of the first text and the dependence relationships between the words of the each second text.

[0040] In a possible implementation mode, in the step S141, dependency grammar dependency grammar analysis may be made over the first text and each second text respectively to obtain the dependence relationships between the multiple words of the first text and the dependence relationships between words of each second text.

[0041] In the example, dependency grammar dependency grammar analysis over the first text may be made to obtain the dependence relationships between the multiple words of the first text.

[0042] Fig. 3 is a schematic diagram illustrating dependence relationships between words of a first text, according to an exemplary embodiment of the present disclosure. As shown in Fig. 3, root may represent a root node of the dependence relationship, and the root node corresponds to the word "很好 (good)"; HED represents that the word "很好 (good)" is a core of the first text, and "很好 (good)" corresponds to "今天 (today)", "天气 (weather)", "," and "今天/北京/故宫/的天气/, /人/很多 (many)" respectively; ADV represents that the dependence relationships between "很好 (good)" and "今天 (today)" is an adverbial-head relationship; SBV represents that the dependence relationships between "很好 (good)" and "天气 (weather)" is a subject-predicate relationship; WP represents that "," is a punctuation mark after "很好 (good)"; COO represents that the dependence relationships between "很好 (good)" and "很多 (many)" is a coordinate relationship, and "天气 (weather)" corresponds to "故宫 (the Forbidden City)"; ATT represents that the relationship between "天气 (weather)" and "故宫 (the Forbidden City)" is an attribute-head relationship, and "故宫 (the Forbidden City)" corresponds to "北京 (beijing)" and "的 (of)" respectively; ATT represents that the dependence relationships between "故宫 (the Forbidden City)" and "北京 (beijing)" is an attribute-head relationship; RAD represents that the relationship between "故宫 (the Forbidden City)" and "的 (of)" is a right auxiliary relationship, and "很多 (many)" corresponds to "人 (people)"; and SBV represents that the relationship between "很多 (many)" and "人 (people)" is the subject-predicate relationship.

[0043] In the example, dependency grammar analysis may be made over the multiple second texts respectively to obtain each dependence relationships between words of each second text. For example, dependency grammar analysis may be made over the second text "北京/故宫/的/天气/很好/, /人/很多 (beijing/the Forbidden City/of/weather/good/, /people/many)" corresponding to "故宫 (today)" to obtain each dependence relationships between the words of the each second text.

[0044] Fig. 4 is a schematic diagram illustrating dependence relationships between words of a second text, according to an exemplary embodiment of the present disclosure. As shown in Fig. 4, root may represent a root node of the dependence relationship, and the root node corresponds to the word "很好 (good)"; HED represents that the word "很好 (good)" is a core word of the first text, and "很好 (good)" corresponds to "天气 (weather)", "," and "很多 (many)" respectively; SBV represents that the dependence relationships between "很好 (good)" and "天气 (weather)" is a subject-predicate relationship; WP represents that "," is a punctuation mark after "很好 (good)"; COO represents that the dependence relationships between "很好 (good)" and "很多 (many)" is a coordinate relationship, and "天气 (weather)" is directed to "故宫 (the Forbidden City)"; ATT represents that the relationship between "天气 (weather)" and "故宫 (the Forbidden City)" is an attribute-head relationship, and "故宫 (the Forbidden City)" corresponds to "北京 (beijing)" and "的 (of)" respectively; ATT represents that the dependence relationships between "故宫 (the Forbidden City)" and "北京 (beijing)" is an attribute-head relationship; RAD represents that the relationship between "故宫 (the Forbidden City)" and "的 (of)" is a right auxiliary relationship, and "很多 (many)" corresponds to "人 (people)"; and SBV represents that the relationship between "很多 (many)" and "人 (people)" is a subject-predicate relationship.

[0045] In the example, based on the dependency grammar analysis processing, dependence relationships between words may also be obtained, including: the dependence relationships between the words of the each second text "今天/北京/故宫/的/天气/很好/, /人/很多 (many)" corresponding to 故宫 the dependence relationships between the words of the each second text "今天/北京/故宫/的/天气/很好/, /人/很多 (today/beijing/the Forbidden City/weather/good/, /people/many)" corresponding to "的 (of)", the dependence relationships between the words of the each second text "今天/北京/故宫/的/天气/很好/, /人/很多 (today/beijing/the Forbidden City/of/good/, /people/many)" corresponding to "人 (weather)", the dependence relationships between the words of the each second text "今天/北京/故宫/的/天气/很好/, /人/很多 (today/beijing/the Forbidden City/of/weather, /people/many)" corresponding to "很多 (good)", the dependence relationships between the words of the each second text "今天/北 京 / 故 宫 / 的 天 气 / 很 好 / 人 / 很 多 (today/beijing/the Forbidden City/of/weather/good/people/many)" corresponding to ",", the dependence relationships between the words of the each second text "今天/北京/故宫/的/天气/很好/, /人/很多 (today/beijing/the Forbidden City/of/weather/good/, /many)" corresponding to "人 (people)". and the dependence relationships between the words of the each second text "今天/北京/故宫/ 的天气/很好/, /人 (today/beijing/the Forbidden City/of/weather/good/, /people)" corresponding to "很好 (many)".

[0046] In a possible implementation mode, in the step S142, the first tree structure of the first text may be obtained according to the dependence relationships between the multiple words of the first text, and the second tree structure of each second text may also be obtained according to the dependence relationships between words of the each second text.

[0047] In the example, the first tree structure may be determined according to the dependence relationships between the multiple words of the first text, namely each word of the first text is taken as a node and the first tree structure is determined according to a correspondence of each word in the dependence relationship.

[0048] Fig. 5 is a schematic diagram illustrating a first tree structure, according to an exemplary embodiment of the present disclosure. As shown in Fig. 5, the core word "很好 (is god)" represented by root is a first node, the node "很好 (is god)" has four child nodes, namely the four words "今天 (today)", "天气 (weather)", "," and "很多 (many)" corresponding to "很好 (is god)" in the dependence relationships of the first text; "今天 (today)" and "," have no child nodes; "天气 (weather)" has a child node "故宫 (the Forbidden City)"; 故宫 (the Forbidden City)" has two child nodes "北京 (beijing)" and "的 (of)"; and "很多 (many)" has a child node "人 (people)".

[0049] In the example, the second tree structure may be determined according to the dependence relationships between the words of the each second text, namely each word of the second text is taken as a node and the second tree structure is determined according to a correspondence of each word in the dependence relationship.

[0050] Fig. 6 is a schematic diagram illustrating a second tree structure, according to an exemplary embodiment of the present disclosure. As shown in Fig. 6, for example, in the second tree structure determined according to the dependence relationships between words of "今天/北京/故宫/的/天气/很好/, /人/很多 (beijing/the Forbidden City/of/weather/good/, /people/many)", the core word "很好 (good)" represented by root is a first node, the node "很好 (good)" has three child nodes, namely the three words "天气 (weather)", "," and "很多 (many)" corresponding to "很好 (good)" in the dependence relationships of the second text; "," has no child nodes; "天气 (weather)" has a child node "故宫 (the Forbidden City) "; "故宫 (the Forbidden City)" has two child nodes "北京 (beijing)" and "的 (of)"; and "很多 (many)" has one child node "A(people)".

[0051] In the example, based on the dependence relationships between words of each second text, different second tree structures may also be obtained, including the second tree structure of "今天/北京/故宫/的/天气/很好/, /人/很多 (today/beijing/ of/weather/good/, /people/many)", the second tree structure of "今天/北京/故宫/的/天气/很好/, /人/很多 (today/beijing/the Forbidden City/weather/good/, /people/many)", the second tree structure of "今天/北京/故宫/的/天气/很好/, /人/很多 (today/beijing/the Forbidden City/of/good/, /people/many)", the second tree structure of "今天/北京/故宫/的/天气/很好/, /人/很多 (today/beijing/the Forbidden City/of/weather/is good /people/many)", the second tree structure of "今天/北京/故宫/的/天气/很好/，/人/很多 (today/beijing/the Forbidden City/of/weather/good/, /many)" and the second tree structure of " 的天气/很好 , /人 (today/beijing/the Forbidden City/of/weather/good/, /people)".

[0052] In such a manner, the first tree structure of the first text and the second tree structures of the second texts may be determined by dependency grammar analysis processing, and information such as word orders and syntactic structures may be reserved in the tree structures, so that the summary text obtained through the first tree structure and the second tree structures can be semantically fluent.

[0053] In a possible implementation mode, in the step S15, a redundant word (i.e., a word that can be deleted with relatively little semantic influence) in the multiple words of the first text may be determined and deleted according to the first tree structure and the multiple second tree structures, and the remaining words may form the summary text of the first text.

[0054] Fig. 7 is a flow chart illustrating the step S15 of a method for text processing, according to an exemplary embodiment of the present disclosure. As shown in Fig. 7, the step S15 includes the following steps.

[0055] In step S151, scores of structure match between multiple second tree structures and the first tree structure are determined respectively according to the first tree structure and each second tree structure.

[0056] In step S152, a target tree structure is determined from the each second tree structure according to the scores of structure match of the multiple second tree structures.

[0057] In step S153, a target word is deleted from the first text to obtain the summary text of the first text, and the target word is a word missing from the second text with the target tree structure relative to the first text.

[0058] In a possible implementation mode, in the step S151, the scores of structure match between the first tree structure and the multiple second tree structures may be determined respectively. In the example, the score of structure match may include a score of structure full match and a score of structure part match.

[0059] In a possible implementation mode, the step S151 may include that: a score of node match between a first target node and a second target node is determined according to the first target node of the first tree structure, a child node of the first target node, the second target node of the second tree structure and a child node of the second target node; the first target node is any word in the first tree structure, and the second target node is a word the same as the first target node; and the scores of structure match between the second tree structures and the first tree structure are determined according to the scores of node match of multiple nodes.

[0060] In an example, the scores of node match of the same words in the first tree structure and the second tree structures may be determined respectively. The score of node match may include a score of node full match and a score of node part match.

[0061] Fig. 8 is a schematic diagram illustrating scores of structure match, according to an exemplary embodiment of the present disclosure. As shown in Fig. 8, the first text is "今天/北京/故宫/的/天气/很好/，/人/很多 (today/beijing/the Forbidden City/ofweather/good/, /people/many)". For example, for the second text "今天/北京/故宫/的/天气/很好/，/人/很多 (beijing/the Forbidden City/of/weather/good/, /people/many)" obtained by removing "今天 (today)", a score of node match between each same word may be determined respectively, namely a score of node full match and a score of node part match of the same word may be determined. In an example, a score of node part match may include a jaccard distance score and a hierarchical discount score.

[0062] In an example, a score of node match between each node (i.e., word) may be determined. For example, a score of node match may be determined according to a child node of each node. For example, the first target node and the second target node are "今天 (today)", the first tree structure has the node "今天 (today)", the second tree structure does not have the node "今天 (today)", and the score of node match of the node "今天 (today)" may be not calculated or the score of node match is determined to be 0. The first target node and the second target node are "北京 (beijing)", the nodes "北京 (beijing)" in both the first tree structure and the second tree structure have no child nodes, and the score of node match of the nodes "北京 " may be not calculated or the score of node match is determined to be 0. The first target node and the second target node are "故宫 (the Forbidden City)", and the nodes "故宫 (the Forbidden City)" in both the first tree structure and the second tree structure have child nodes "北京 (beijing)" and "的 (of)", so that the sets formed by the child nodes of "故宫 (the Forbidden City)" in both the first tree structure and the second tree structure are the same, namely the set formed by the child nodes of "故宫 (the Forbidden City)" in the first tree structure is {"北京 (beijing)", "的 (of)"} and includes two elements, and the set formed by the child nodes of "故宫 (the Forbidden City)" in the second tree structure is {"北京 (beijing)", "的 (of)"} and includes two elements. A score of node full match of the two sets is calculated through a full match algorithm, and the score of node full match is 1 (the score of node full match is 1 when the child nodes are completely the same, and is 0 when the child nodes are not completely the same). When a score of node part match is calculated, a jaccard distance of the two sets may be calculated to obtain that a jaccard distance score is 1, and a hierarchical discount score of the two sets may be calculated to be 1 (a discount rate may be set to be 0.5, namely the hierarchical discount score is 1 if the child nodes are the same, while the hierarchical discount score is 0.5 if the child nodes are different but child nodes of the child nodes are the same, and the hierarchical discount score is multiplied by 0.5 if the child nodes of the child nodes are different). In such a calculation manner, scores of node match of "的 (of)", "," and "人 (people)" may be determined to be 0, and scores of node match of "天气 (weather)" and "很多 (many)" may both be determined to be 1. "很好 (good)" in the first tree structure has four child nodes, i.e., "今天 (today)", "天气 (weather)", "," and "很多 (many)"; a set formed by the child nodes of "很好 (good)" in the first tree structure is {"今天 (today)", "天气 (weather)", ",", "很多 (many)"} and includes four elements; and a set formed by the child nodes of "很好 (good)" in the second tree structure is {"天气 (weather)", ",", "很多 (many)"} and includes three elements. A score of node full match between the two sets is calculated through the full match algorithm, and the score of node full match is 0. When a score of node part match is calculated, a jaccard distance between the two sets may be calculated to obtain that a jaccard distance score is 0.75, and a hierarchical discount score of the two sets may be calculated to be 0.5.

[0063] In a possible implementation mode, the score of structure match between the first tree structure and the second tree structure may be determined according to the scores of node match of all the nodes. In the example, the scores of node full match of all the nodes may be summed to obtain the score of structure full match, and the scores of node part match of all the nodes may be summed to obtain the score of structure part match.

[0064] In an example, the first text is "今天/北京/故宫/的/天气/很好/, /人/很多 (today/beijing/the Forbidden City/of/weather/good/, /people/many)", the second text is "北京 / 今天/北京/故宫/的/天气/很好/, /人/很多 (beijing/the Forbidden City/of/weather/good/, /people/many)", the score of structure full match is 3, and for the score of structure part match, a total jaccard distance score is 3.75 and a total hierarchical discount score is 3.5.

[0065] In a possible implementation mode, the score of structure full match and score of structure part match between the first tree structure and each second tree structure are determined respectively in the aforesaid manner.

[0066] In a possible implementation mode, in the step S152, a target tree structure may be determined according to the scores of structure match of the multiple second tree structures. In an example, the scores of structure match between the first tree structure and the second tree structures may be arranged in a descending order. In a process of arrangement, arrangement may be performed according to the scores of structure full match at first; if the scores of structure full match between two or more second tree structures and the first tree structure are the same, the two or more second tree structures are arranged according to the scores of structure part match. In a process of arrangement according to the scores of structure part match, arrangement in a descending order may be performed according to the jaccard distance scores at first, and if the jaccard distance scores are the same, arrangement in a descending order may be performed according to the hierarchical discount scores. For example, if a score of structure full match of a second tree structure corresponding to word A is 3 and a score of structure full match of a second tree structure corresponding to word B is 2, then the second tree structure corresponding to the word A is arranged before the second tree structure corresponding to the word B. For another example, upon when the score of structure full match of the second tree structure corresponding to the word A is 3 and the score of structure full match of the second tree structure corresponding to the word B is 3, and further upon that a jaccard distance score of the second tree structure corresponding to the word A is 3.75 and that a jaccard distance score of the second tree structure corresponding to the word B is 4, the second tree structure corresponding to the word B is arranged before the second tree structure corresponding to the word A. For another example, upon when the score of structure full match of the second tree structure corresponding to the word A is 3 and the score of structure full match of the second tree structure corresponding to the word B is 3, and further upon that the jaccard distance score of the second tree structure corresponding to the word A is 3.75 and that the jaccard distance score of the second tree structure corresponding to the word B is 3.75, and still further upon that a hierarchical discount score of the second tree structure corresponding to the word A is 3.5 and that a hierarchical discount score of the second tree structure corresponding to the word B is 3, the second tree structure corresponding to the word A is arranged before the second tree structure corresponding to the word B. Thus, an order of the multiple second tree structures can be obtained through the aforesaid method.

[0067] In the example, the first i second tree structures may be selected as target tree structures, i (i<n) being a preset value. The scores of structure match between the first i second tree structures and the first tree structure are relatively high, namely semantic similarities between the second texts corresponding to the first i second tree structures and the first text are relatively high, so that words missing from the second texts with the first i second tree structures relative to the first text have relatively little semantic influence on the first text.

[0068] In a possible implementation mode, in the step S153, the words (i.e., target words) missing from the second text with the first i second tree structures relative to the first text may be deleted from the first text, namely a summary text with a higher semantic similarity with the first text may be obtained under the condition of having less semantic influence on the first text.

[0069] In such a manner, a target word may be determined according to the score of structure match between the first tree structure and the second tree structure to obtain the summary text, so that a redundant word may be removed to obtain the summary text with a relatively high semantic similarity with the first text under the condition of having less semantic influence on the first text.

[0070] Fig. 9 is a schematic diagram illustrating application of a method for text processing, according to an exemplary embodiment of the present disclosure. As shown in Fig. 9, a word segmentation result of a first text "今天/北京/故宫/的/天气/很好/，/人/很多(Weather of the Forbidden City in Beijing is good, many people)" is "今天/北京/故宫/的/天气/很好/，/人/很多 (today/beijing/the Forbidden City/of/weather/good/, /people/many)".

[0071] In a possible implementation mode, each word may be removed from the first text to obtain nine second texts, i.e., "今天/北京/故宫/的/天气/很好/，/人/很多 (beijing/the Forbidden City/of/weather/good/, /people/many)", "今天/北京/故宫/的/天气/很好/，/人/很多 (today/the Forbidden City/of/weather/good/, /people/many)", "今天/北京/故宫/的/天气/很好/，/人/很多 (today/beijing/of/weather/good/, /people/many)", "今天/北京/故宫/的/天气/很好/，/人/很多 (today/beijing/the Forbidden City/weather/good/, /people/many)", 今天/北京/故宫/的/天气/很好/，/人/很多 (today/beijing/the Forbidden City/of/good/, /people/many)", "今天/北京/故宫/的/天气/很好/，/人/很多 (today/beijing/the Forbidden City/of/weather, /people/many)", "今天/北京/故宫/的/天气/很好/，/人/很多 (today/beijing/the Forbidden City/of/weather/is good /people/many)", "今天/北京/故宫/的/天气/很好/，/人/很多 (today/beijing/the Forbidden City/of/weather/good/, /many)" and "今天/北京/故宫/的天气/很好 /，/人 (today/beijing/the Forbidden City/of/weather/good/, /people)".

[0072] In a possible implementation mode, dependency grammar analysis may be made over the first text and the nine second texts to obtain dependence relationships between the multiple words of the first text and dependence relationships between words of each second text, and a first tree structure and nine second tree structures are determined according to the dependence relationships between all the words, for example, correspondences between words in the dependence relationship.

[0073] In a possible implementation mode, a total jaccard distance score, a total hierarchical discount score and a score of structure full match between the first tree structure and each second tree structure may be determined respectively, and arrangement in a descending order is arranged according to the scores of structure full match, the total jaccard distance scores and the total hierarchical discount scores. In an example, arrangement may be performed according to scores of structure full match at first; under the condition that the scores of structure full match are the same arrangement is performed according to the total jaccard distance scores; and under the condition that the total jaccard distance scores are the same, arrangement is performed according to hierarchical discount scores. In such a manner, an order of the nine second tree structures may be obtained.

[0074] In a possible implementation mode, the first two second tree structures may be determined as target tree structures, and words corresponding to the first two second tree structures are determined as target words, the target words being two words with minimum semantic influence on the first text. In an example, the target words are "的 (of)" and ",", and "的 (of)" and "," may be removed from the first text to obtain "今天/北京/故宫/的/天气/很好/，/人/很多 (today/beijing/the Forbidden City/weather/good/people/many)", namely a summary text of the first text is "今天北京故宫天气很好人很多 (weather the Forbidden City in Beijing is good many people)".

[0075] According to the method for text processing of the embodiment of the present disclosure, manual annotation and training are avoided, and a workload in manual annotation is reduced. Since training of a large amount of annotation data is not needed, the range of application can be extended. A second text corresponding to each word of the first text may be obtained by deleting each word from the first text, so that a probability of ignoring key information is reduced. The first tree structure of the first text and the second tree structures of the second texts may be determined by dependency grammar analysis, and information such as word orders and syntactic structures may be reserved in the tree structures, so that the summary text obtained through the first tree structure and the second tree structures is semantically fluent. Furthermore, the target word may be determined according to the score of structure match between the first tree structure and the second tree structure to obtain the summary text, so that the summary text with the relatively high semantic similarity with the first text may be obtained by removing a redundant word under the condition of having little semantic influence on the first text.

[0076] Fig. 10 is a block diagram of a device for text processing, according to an exemplary embodiment of the present disclosure. As shown in Fig. 10, the device includes:

a first module 11, configured to perform word segmentation on a first text to be processed to determine multiple words of the first text;
a second module 12, configured to remove, from the first text, one word in the multiple words of the first text to obtain one second text;
a third module 13, configured to remove each word from the first text one by one to obtain a second text set;
a fourth module 14, configured to analyze the first text and each second text in the second text set respectively to obtain a first tree structure of the first text and a second tree structure of each second text, each word in the first text being a node of the first tree structure and each word in the second text being a node of the second tree structure; and
a fifth module 15, configured to determine a summary text of the first text according to the first tree structure and each second tree structure.

[0077] Fig. 11 is a block diagram of a device for text processing, according to an exemplary embodiment of the present disclosure. As shown in Fig. 11, the fourth module 14 includes:

a first submodule 141, configured to make dependency grammar analysis over the first text and each second text respectively to obtain dependence relationships between the multiple words of the first text and dependence relationships between words of each second text respectively; and
a second submodule 142, configured to obtain the first tree structure of the first text and the second tree structure of each second text respectively according to the dependence relationships between the multiple words of the first text and the dependence relationships between the words of the each second text.

[0078] In a possible implementation mode, the fifth module 15 may include:

a third submodule 151, configured to determine scores of structure match between multiple second tree structures and the first tree structure respectively according to the first tree structure and the each second tree structure;
a fourth submodule 152, configured to determine a target tree structure from the each second tree structure according to the scores of structure match of the second tree structures; and
a fifth submodule 153, configured to delete a target word from the first text to obtain the summary text of the first text, the target word being a word missing from the second text with the target tree structure relative to the first text.

[0079] In a possible implementation mode, the third submodule 151 is further configured to:

determine a score of node match between a first target node and a second target node according to the first target node of the first tree structure, a child node of the first target node, the second target node of the second tree structure and a child node of the second target node, the first target node being any word in the first tree structure and the second target node being a word the same as the first target node; and determine scores of structure match between the second tree structures and the first tree structure according to the scores of node match of multiple nodes.

[0080] In a possible implementation mode, the score of structure match may include a score of structure full match and a score of structure part match.

[0081] With respect to the device in the above embodiment, the specific way for performing operations by individual modules therein has been described in detail in the embodiment regarding the method, which will not be elaborated herein.

[0082] Fig. 12 is a block diagram of a device 800 for text processing, according to an exemplary embodiment of the present disclosure. For example, the device 800 may be a mobile phone, a computer, a digital broadcast terminal, a messaging device, a gaming console, a tablet, a medical device, exercise equipment, a personal digital assistant and the like.

[0083] Referring to Fig. 12, the device 800 may include one or more of the following components: a processing component 802, a memory 804, a power component 806, a multimedia component 808, an audio component 810, an Input/Output (I/O) interface 812, a sensor component 814, and a communication component 816.

[0084] The processing component 802 is typically configured to control overall operations of the device 800, such as the operations associated with display, telephone calls, data communications, camera operations, and recording operations. The processing component 802 may include one or more processors 820 to execute instructions to perform all or part of the steps in the aforesaid method. Moreover, the processing component 802 may include one or more modules which facilitate interaction between the processing component 802 and the other components. For instance, the processing component 802 may include a multimedia module to facilitate interaction between the multimedia component 808 and the processing component 802.

[0085] The memory 804 is configured to store various types of data to support the operation of the device 800. Examples of such data include instructions for any application programs or methods operated on the device 800, contact data, phonebook data, messages, pictures, video, etc. The memory 804 may be implemented by any type of volatile or non-volatile memory devices, or a combination thereof, such as a Static Random Access Memory (SRAM), an Electrically Erasable Programmable Read-Only Memory (EEPROM), an Erasable Programmable Read-Only Memory (EPROM), a Programmable Read-Only Memory (PROM), a Read-Only Memory (ROM), a magnetic memory, a flash memory, and a magnetic or optical disk.

[0086] The power component 806 is configured to provide power for various components of the device 800. The power component 806 may include a power management system, one or more power supplies, and other components associated with generation, management and distribution of power for the device 800.

[0087] The multimedia component 808 may include a screen for providing an output interface between the device 800 and a user. In some embodiments, the screen may include a Liquid Crystal Display (LCD) and a Touch Panel (TP). If the screen includes the TP, the screen may be implemented as a touch screen to receive an input signal from the user. The TP may include one or more touch sensors to sense touches, swipes and gestures on the TP. The touch sensors may not only sense a boundary of a touch or swipe action but also detect a duration and pressure associated with the touch or swipe action. In some embodiments, the multimedia component 808 may include a front camera and/or a rear camera. The front camera and/or the rear camera may receive external multimedia data when the device 800 is in an operation mode, such as a photographing mode or a video mode. Each of the front camera and the rear camera may be a fixed optical lens system or have focusing and optical zooming capabilities.

[0088] The audio component 810 is configured to output and/or input an audio signal. For example, the audio component 810 may include a Microphone (MIC), and the MIC is configured to receive an external audio signal when the device 800 is in the operation mode, such as a call mode, a recording mode and a voice recognition mode. The received audio signal may further be stored in the memory 804 or sent through the communication component 816. In some embodiments, the audio component 810 may further include a speaker configured to output the audio signal.

[0089] The I/O interface 812 is configured to provide an interface between the processing component 802 and a peripheral interface module, and the peripheral interface module may be a keyboard, a click wheel, a button and the like. The button may include, but not limited to: a home button, a volume button, a starting button and a locking button.

[0090] The sensor component 814 may include one or more sensors configured to provide status assessment in various aspects for the device 800. For instance, the sensor component 814 may detect an on/off status of the device 800 and relative positioning of components, such as a display and small keyboard of the device 800, and the sensor component 814 may further detect a change in a position of the device 800 or a component of the device 800, presence or absence of contact between the user and the device 800, orientation or acceleration/deceleration of the device 800 and a change in temperature of the device 800. The sensor component 814 may include a proximity sensor configured to detect presence of an object nearby without any physical contact. The sensor component 814 may also include a light sensor, such as a Complementary Metal Oxide Semiconductor (CMOS) or Charge Coupled Device (CCD) image sensor, configured for use in an imaging APP. In some embodiments, the sensor component 814 may also include an acceleration sensor, a gyroscope sensor, a magnetic sensor, a pressure sensor or a temperature sensor.

[0091] The communication component 816 is configured to facilitate wired or wireless communication between the device 800 and other equipment. The device 800 may access a communication-standard-based wireless network, such as a Wireless Fidelity (WiFi) network, a 2nd-Generation (2G) or 3rd-Generation (3G) network or a combination thereof. In an exemplary embodiment, the communication component 816 is configured to receive a broadcast signal or broadcast associated information from an external broadcast management system through a broadcast channel. In an exemplary embodiment, the communication component 816 may further include a Near Field Communication (NFC) module to facilitate short-range communication. For example, the NFC module may be implemented based on a Radio Frequency Identification (RFID) technology, an Infrared Data Association (IrDA) technology, an Ultra-WideBand (UWB) technology, a Bluetooth (BT) technology and another technology.

[0092] In an exemplary embodiment, the device 800 may be implemented by one or more Application Specific Integrated Circuits (ASICs), Digital Signal Processors (DSPs), Digital Signal Processing Devices (DSPDs), Programmable Logic Devices (PLDs), Field Programmable Gate Arrays (FPGAs), controllers, micro-controllers, microprocessors or other electronic components, and is configured to execute the aforesaid method.

[0093] In an exemplary embodiment, there is also provided a non-transitory computer-readable storage medium including an instruction, such as the memory 804 including an instruction, and the instruction may be executed by the processor 820 of the device 800 to implement the aforesaid method. For example, the non-transitory computer-readable storage medium may be a ROM, a Random Access Memory (RAM), a Compact Disc Read-Only Memory (CD-ROM), a magnetic tape, a floppy disc, an optical data storage device and the like.

[0094] Fig. 13 is a block diagram of a device 1900 for text processing, according to an exemplary embodiment of the present disclosure. For example, the device 1900 may be provided as a server. Referring to Fig. 13, the device 1900 includes a processing component 1922 which further includes one or more processors, and a memory resource represented by a memory 1932 configured to store an instruction executable by the processing component 1922, for example, an application program. The application program stored in the memory 1932 may include one or more than one module of which each corresponds to a set of instructions. In addition, the processing component 1922 is configured to execute the instruction to execute the aforesaid method.

[0095] The device 1900 may further include a power component 1926 configured to execute power management of the device 1900, a wired or wireless network interface 1950 configured to connect the device 1900 to a network, and an I/O interface 1958. The device 1900 may be operated based on an operating system stored in the memory 1932, for example, Windows ServerTM, Max OS XTM, UnixTM, LinuxTM, FreeBSDTM or the like.

[0096] In an exemplary embodiment, there is also provided a non-transitory computer-readable storage medium including an instruction, such as the memory 1932 including an instruction, and the instruction may be executed by the processing component 1922 of the device 1900 to implement the aforesaid method. For example, the non-transitory computer-readable storage medium may be a ROM, a RAM, a CD-ROM, a magnetic tape, a floppy disc, an optical data storage device and the like.

[0097] Other implementation solutions of the present disclosure will be apparent to those skilled in the art from consideration of the specification and practice of the present disclosure. This application is intended to cover any variations, uses, or adaptations of the present disclosure following the general principles thereof and including such departures from the present disclosure as come within known or customary practice in the art. It is intended that the specification and examples be considered as exemplary only, with a true scope of the present disclosure being indicated by the following claims.

[0098] It will be appreciated that the present disclosure is not limited to the exact construction that has been described above and illustrated in the accompanying drawings, and that various modifications and changes may be made without departing from the scope thereof. It is intended that the scope of the present disclosure only be limited by the appended claims.

Claims

1. A method for text processing, **characterized in that**, comprising:

performing (S11) word segmentation on a first text to be processed to determine multiple words of the first text;
removing (S12), from the first text, one word in the multiple words of the first text to obtain one second text;
removing (S13) each word from the first text one by one to obtain a second text set;
analyzing (S14) the first text and each second text in the second text set respectively to obtain a first tree structure of the first text and a second tree structure of the each second text, wherein each of the multiple words in the first text is a node of the first tree structure, and each of words in the each second text is a node of the each second tree structure; and
determining (S15) a summary text of the first text according to the first tree structure and the each second tree

structure.

2. The method of claim 1, wherein analyzing the first text and the each second text in the second text set respectively to obtain the first tree structure of the first text and the second tree structure of the each second text comprises:

   making dependency grammar analysis over the first text and the each second text respectively to obtain dependence relationships between the multiple words of the first text and dependence relationships between the words of the each second text respectively; and
   obtaining the first tree structure of the first text and the second tree structure of the each second text respectively according to the dependence relationships between the multiple words of the first text and the dependence relationships between the words of the each second text.

3. The method of claim 1, wherein determining the summary of the text to be processed according to the first tree structure and the each second tree structure comprises:

   determining scores of structure match between multiple second tree structures and the first tree structure respectively according to the first tree structure and the each second tree structure;
   determining a target tree structure from the each second tree structure according to the scores of structure match of the multiple second tree structures; and
   deleting a target word from the first text to obtain the summary text of the first text, wherein the target word is a word missing from a second text with the target tree structure relative to the first text.

4. The method of claim 3, wherein determining the scores of structure match between the multiple second tree structures and the first tree structure respectively according to the first tree structure and the each second tree structure comprises:

   determining a score of node match between a first target node and a second target node according to the first target node of the first tree structure, a child node of the first target node, the second target node of the second tree structure and a child node of the second target node, wherein the first target node is any word in the first tree structure and the second target node is a word the same as the first target node; and
   determining the scores of structure match between the multiple second tree structures and the first tree structure according to the scores of node match of multiple nodes.

5. The method of claim 3 or 4, wherein the scores of structure match comprise scores of structure full match and scores of structure part match.

6. A device for text processing, **characterized in that**, comprising:

   a first module (11), configured to perform word segmentation on a first text to be processed to determine multiple words of the first text;
   a second module (12), configured to remove, from the first text, one word in the multiple words of the first text to obtain one second text;
   a third module (13), configured to remove each word from the first text one by one to obtain a second text set;
   a fourth module (14), configured to analyze the first text and each second text in the second text set respectively to obtain a first tree structure of the first text and a second tree structure of the each second text, wherein each of the multiple words in the first text is a node of the first tree structure, and each of words in the second text is a node of the second tree structure; and
   a fifth module (15), configured to determine a summary text of the first text according to the first tree structure and the each second tree structure.

7. The device of claim 6, wherein the fourth module comprises:

   a first submodule (141), configured to make dependency grammar analysis over the first text and the each second text respectively to obtain dependence relationships between the multiple words of the first text and dependence relationships between the words of the each second text respectively; and
   a second submodule (142), configured to obtain the first tree structure of the first text and the second tree structure of the each second text respectively according to the dependence relationships between the multiple words of the first text and the dependence relationships between the words of the each second text.

8. The device of claim 6, wherein the fifth module comprises:

a third submodule (151), configured to determine scores of structure match between multiple second tree structures and the first tree structure respectively according to the first tree structure and the each second tree structure;
a fourth submodule (152), configured to determine a target tree structure from the each second tree structure according to the scores of structure match of the multiple second tree structures; and
a fifth submodule (153), configured to delete a target word from the first text to obtain the summary text of the first text, wherein the target word is a word missing from a second text with the target tree structure relative to the first text.

9. The device of claim 8, wherein the third submodule is further configured to:

determine a score of node match between a first target node and a second target node according to the first target node of the first tree structure, a child node of the first target node, the second target node of the second tree structure and a child node of the second target node, wherein the first target node is any word in the first tree structure, and the second target node is a word the same as the first target node; and
determine the scores of structure match between the multiple second tree structures and the first tree structure according to the scores of node match of multiple nodes.

10. The device of claim 8 or 9, wherein the scores of structure match comprise scores of structure full match and scores of structure part match.

11. A device for text processing, **characterized in that**, comprising:

a processor (820); and
a memory (804) configured to store an instruction executable for the processor,
wherein the processor is configured to execute the method of any one of claims 1-5.

12. A non-transitory computer-readable storage medium, having instructions stored thereon for execution by a processor to enable the processor to implement the method of any one of claims 1-5.

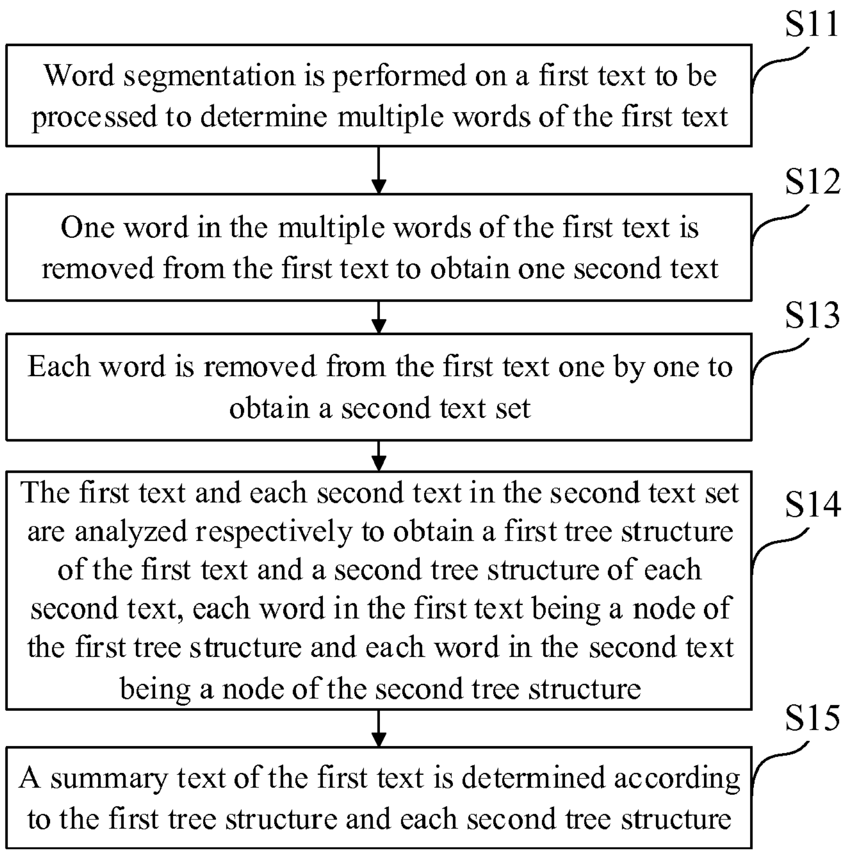
S11
Word segmentation is performed on a first text to be processed to determine multiple words of the first text
S12
One word in the multiple words of the first text is removed from the first text to obtain one second text
S13
Each word is removed from the first text one by one to obtain a second text set
S14
The first text and each second text in the second text set are analyzed respectively to obtain a first tree structure of the first text and a second tree structure of each second text, each word in the first text being a node of the first tree structure and each word in the second text being a node of the second tree structure
S15
A summary text of the first text is determined according to the first tree structure and each second tree structure

FIG. 1

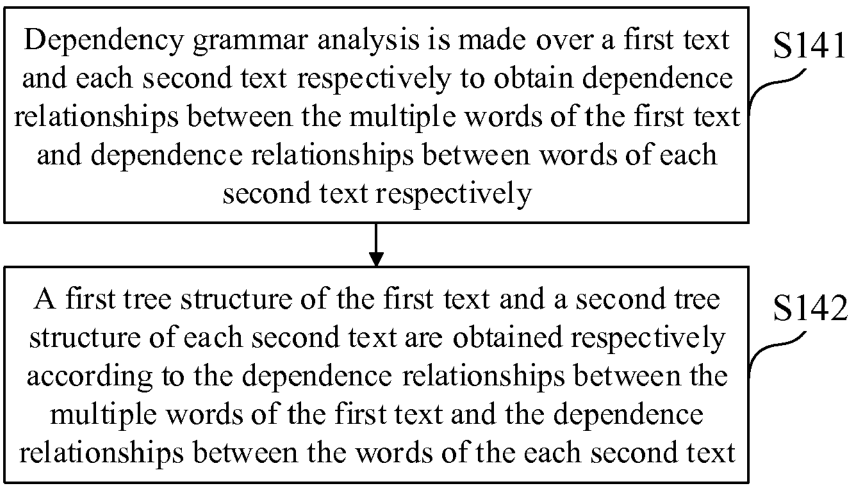
Dependency grammar analysis is made over a first text and each second text respectively to obtain dependence relationships between the multiple words of the first text and dependence relationships between words of each second text respectively
S141
A first tree structure of the first text and a second tree structure of each second text are obtained respectively according to the dependence relationships between the multiple words of the first text and the dependence relationships between the words of the each second text
S142

FIG. 2

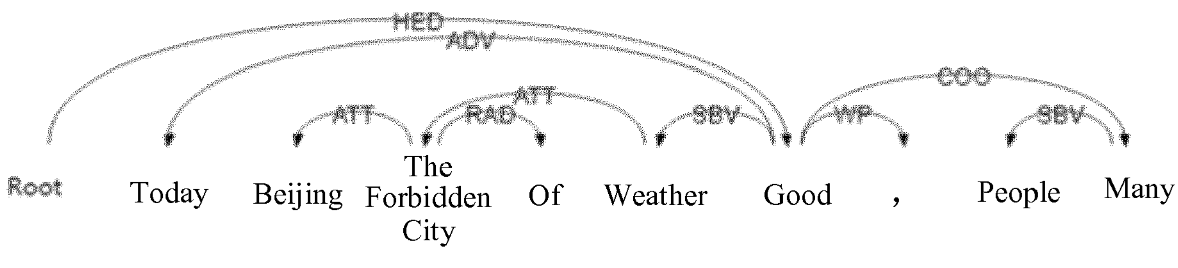
HED
ADV
ATT
RAD
ATT
SBV
COO
WP
SBV
Root
Today
Beijing
The Forbidden City
Of
Weather
Good
,
People
Many

FIG. 3

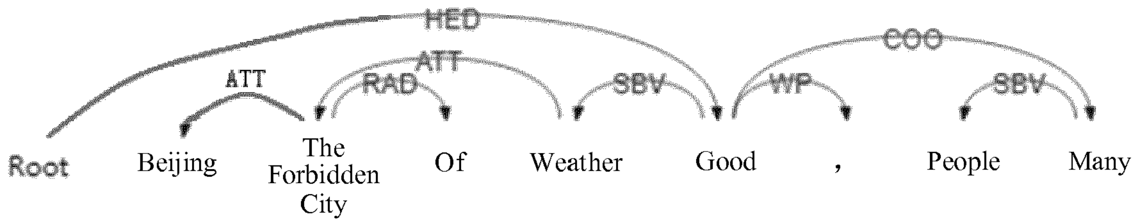

FIG. 4

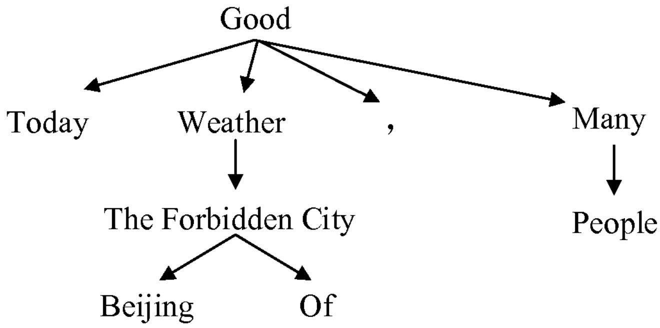

FIG. 5

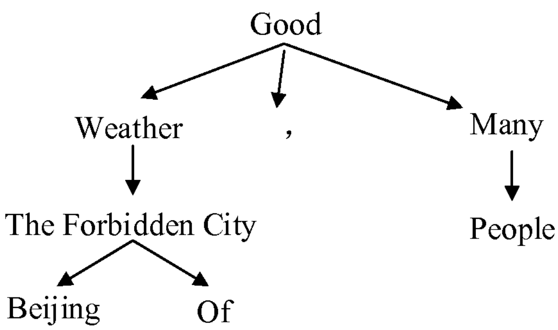

FIG. 6

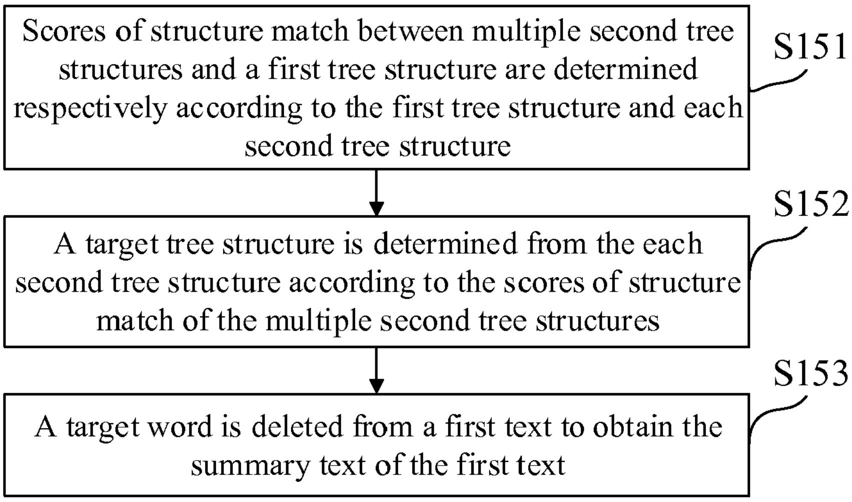

FIG. 7

Word	Child node (first text)	Child node (second text)	Score of node full match	Score of node part match	
				Jaccard distance score	Hierarchical discount score
Today					
Beijing					
The Forbidden City	Beijing/Of	Beijing/Of	1	1	1
Of					
Weather	The Forbidden City	The Forbidden City	1	1	1
Good	Today/Weather/,/Many	Today/,/Many	0	0. 75	0. 5
,					
People					
Many	People	People	1	1	1
Total			3	3. 75	3. 5

FIG. 8

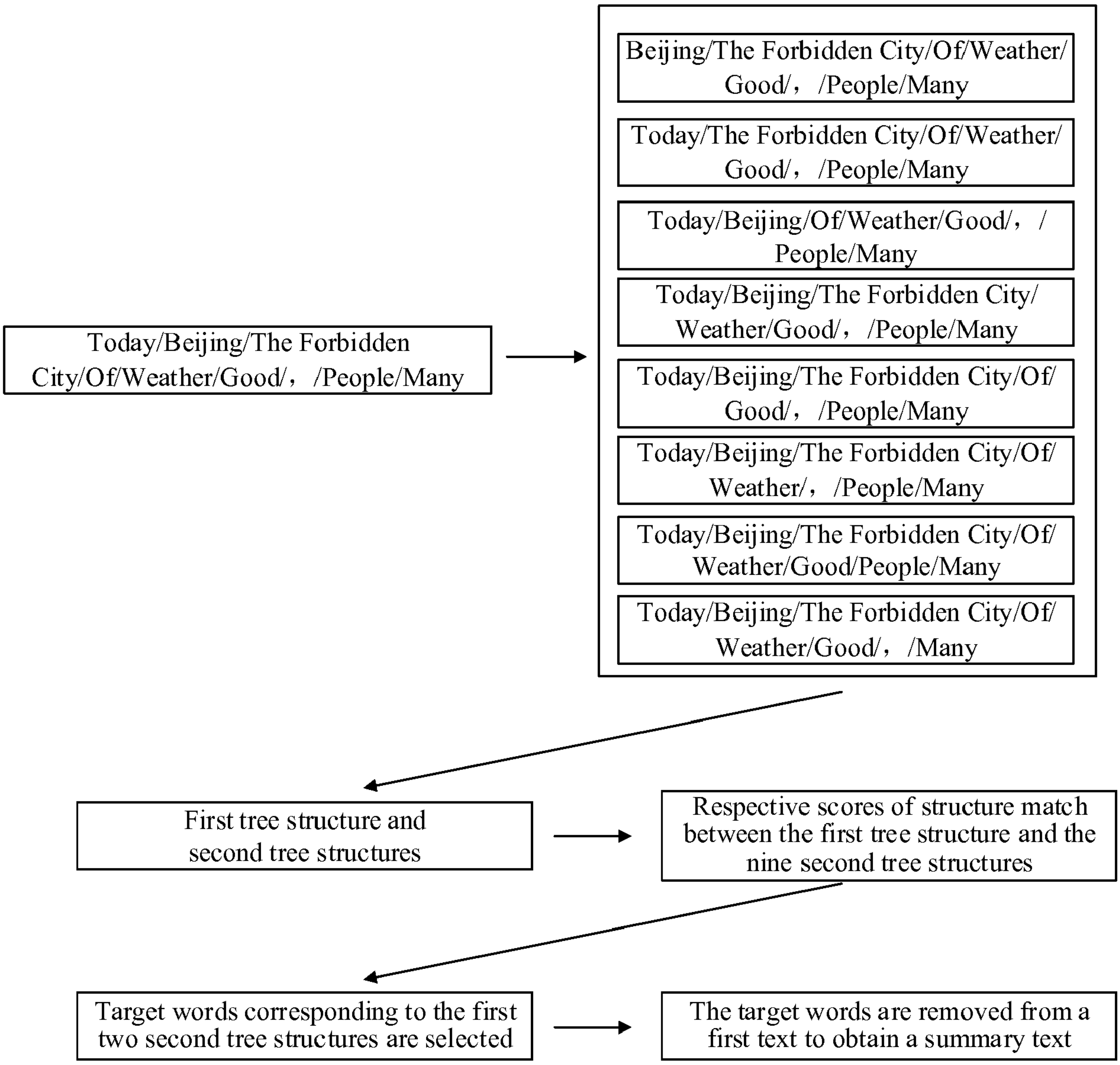
Today/Beijing/The Forbidden City/Of/Weather/Good/，/People/Many
Beijing/The Forbidden City/Of/Weather/Good/，/People/Many
Today/The Forbidden City/Of/Weather/Good/，/People/Many
Today/Beijing/Of/Weather/Good/，/People/Many
Today/Beijing/The Forbidden City/Weather/Good/，/People/Many
Today/Beijing/The Forbidden City/Of/Good/，/People/Many
Today/Beijing/The Forbidden City/Of/Weather/，/People/Many
Today/Beijing/The Forbidden City/Of/Weather/Good/People/Many
Today/Beijing/The Forbidden City/Of/Weather/Good/，/Many
First tree structure and second tree structures
Respective scores of structure match between the first tree structure and the nine second tree structures
Target words corresponding to the first two second tree structures are selected
The target words are removed from a first text to obtain a summary text

FIG. 9

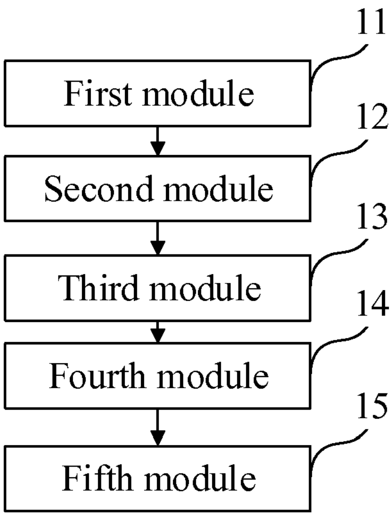
11
First module
12
Second module
13
Third module
14
Fourth module
15
Fifth module

FIG. 10

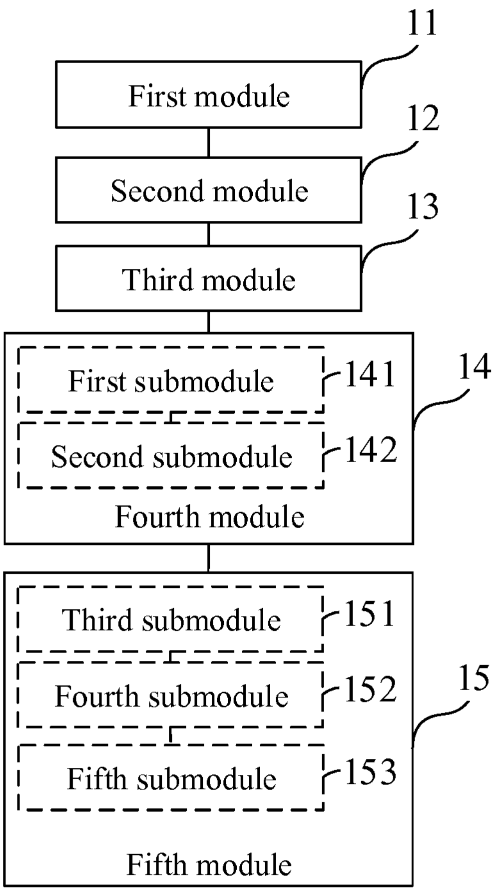
11
First module
12
Second module
13
Third module
First submodule
141
14
Second submodule
142
Fourth module
Third submodule
151
Fourth submodule
152
15
Fifth submodule
153
Fifth module

FIG. 11

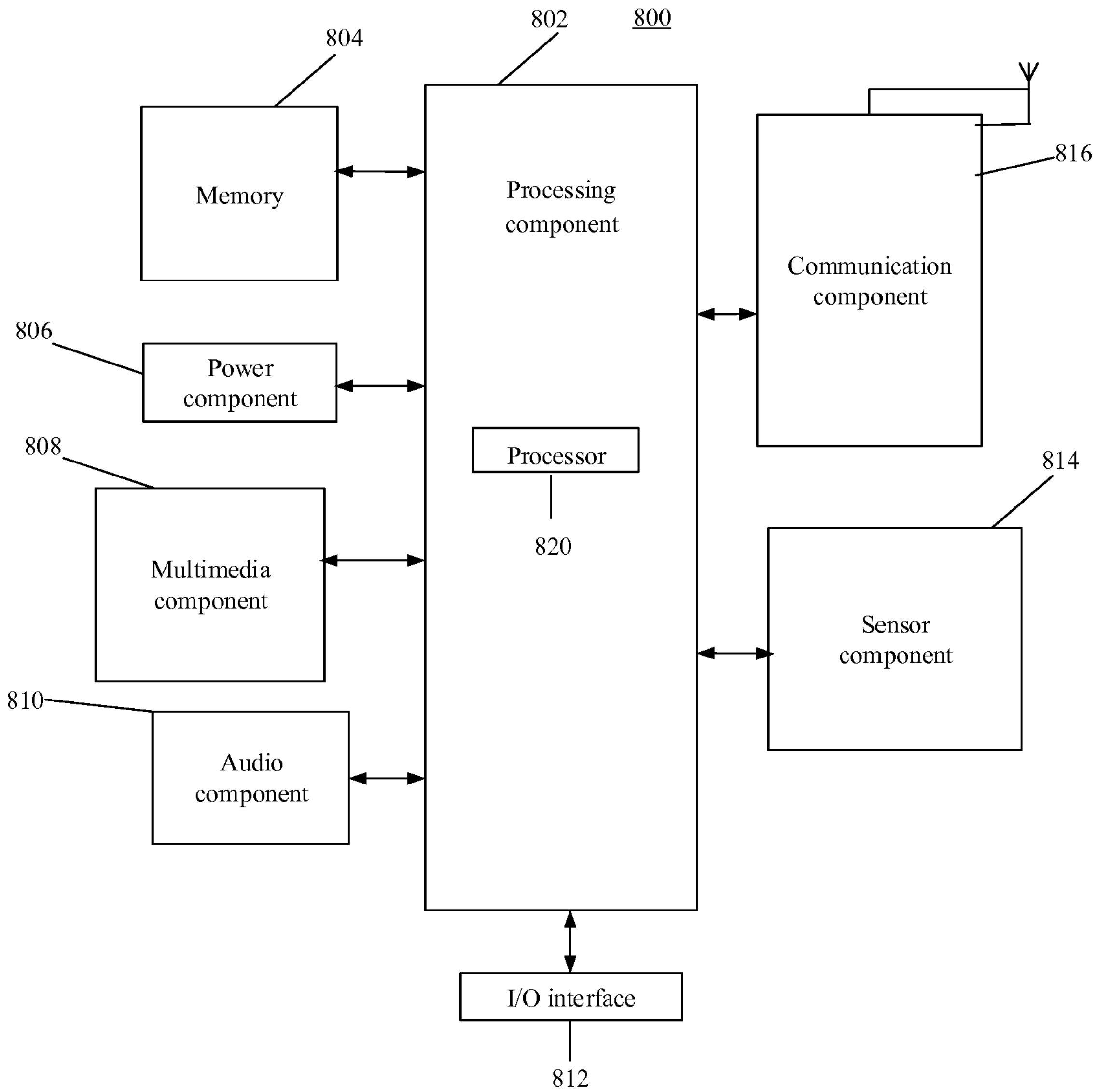
800
802
804
Memory
Processing
component
806
Power
component
Communication
component
816
808
Processor
820
814
Multimedia
component
Sensor
component
810
Audio
component
I/O interface
812

FIG. 12

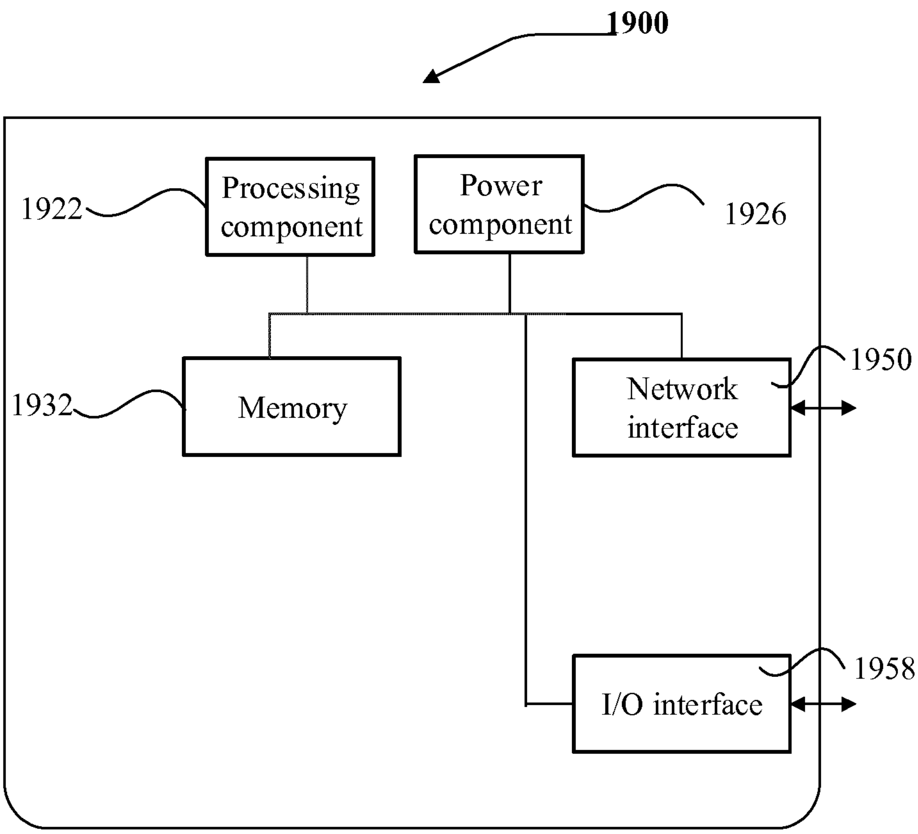
1900
1922
Processing component
Power component
1926
1932
Memory
Network interface
1950
I/O interface
1958

FIG. 13

Europäisches Patentamt
European Patent Office
Office européen des brevets

EUROPEAN SEARCH REPORT

Application Number
EP 19 21 1879

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X	US 2004/044519 A1 (POLANYI LIVIA [US] ET AL) 4 March 2004 (2004-03-04) * the whole document * -----	1-12	INV. G06F16/34 G06F40/30
X	WO 01/86489 A2 (UNIV SOUTHERN CALIFORNIA [US]; MARCU DANIEL [US]; KNIGHT KEVIN [US]) 15 November 2001 (2001-11-15) * page 6, line 5 - line 12 * * page 11, line 6 - page 12, line 18 * * page 41, line 1 - page 43, line 8 * * page 62, line 9 - page 63, line 19; figure 18 * -----	1-12	
X	WO 2009/131417 A2 (KOREA ADVANCED INST SCI & TECH [KR]; CHOI KEY-SUN [KR] ET AL.) 29 October 2009 (2009-10-29) * the whole document * -----	1-12	
X	US 2016/283588 A1 (KATAE NOBUYUKI [JP]) 29 September 2016 (2016-09-29) * the whole document * -----	1-12	TECHNICAL FIELDS SEARCHED (IPC)
X	Saziye Betül Ozates ET AL: "Sentence Similarity based on Dependency Tree Kernels for Multi-document Summarization", , 1 January 2016 (2016-01-01), XP055696978, Retrieved from the Internet: URL:https://www.aclweb.org/anthology/L16-1452.pdf [retrieved on 2020-05-19] * the whole document * -----	1-12	G06F

The present search report has been drawn up for all claims

1

Place of search	Date of completion of the search	Examiner
Berlin	20 May 2020	Lechenne, Laurence

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons
& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 19 21 1879

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report.
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

20-05-2020

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2004044519 A1	04-03-2004	JP 2004094946 A	25-03-2004
		US 2004044519 A1	04-03-2004
WO 0186489 A2	15-11-2001	AU 6150501 A	20-11-2001
		AU 6150601 A	20-11-2001
		CA 2408819 A1	15-11-2001
		CN 1465018 A	31-12-2003
		EP 1352338 A2	15-10-2003
		JP 2004501429 A	15-01-2004
		US 2002040292 A1	04-04-2002
		US 2002046018 A1	18-04-2002
		WO 0186489 A2	15-11-2001
		WO 0186491 A2	15-11-2001
WO 2009131417 A2	29-10-2009	KR 100921563 B1	12-10-2009
		WO 2009131417 A2	29-10-2009
US 2016283588 A1	29-09-2016	NONE	

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82